

(19)



Europäisches Patentamt

European Patent Office

Office européen des brevets



(11) Publication number:

0 552 770 A2

(12)

EUROPEAN PATENT APPLICATION

(21) Application number: 93100900.5

(51) Int. Cl.⁵: G06K 9/46, A61B 5/117

(22) Date of filing: 21.01.93

(30) Priority: 23.01.92 JP 9753/92

(43) Date of publication of application:
28.07.93 Bulletin 93/30(84) Designated Contracting States:
DE FR GB(71) Applicant: **MATSUSHITA ELECTRIC
INDUSTRIAL CO., LTD.**
1006, Oaza Kadoma
Kadoma-shi, Osaka-fu, 571(JP)(72) Inventor: Kado, Yoshiyasu
Showa-ryo, 16-1, Miyamae-cho

Kadoma City, 571(JP)

Inventor: Nakagawa, Masamichi

D405, 1-9-4, Nasuzukuri

Hirakata City, 573(JP)

Inventor: Maehara, Fumio

52-12, Oeda Higashi-cho

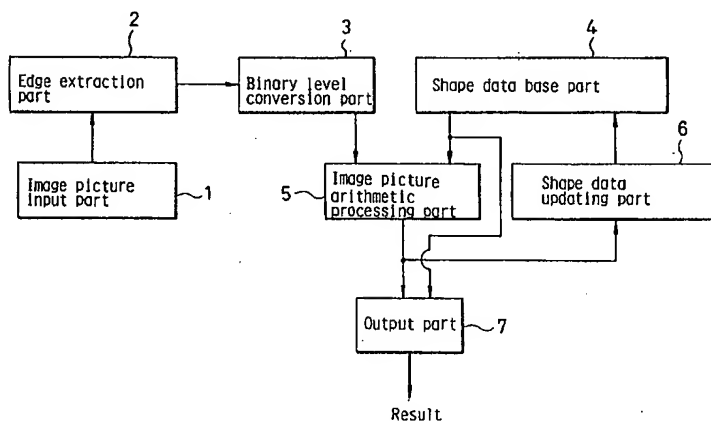
Moriguchi City, 570(JP)

(74) Representative: Jung, Elisabeth, Dr. et al
Dr. Elisabeth Jung Dr. Jürgen Schirdewahn
Dipl.-Ing. Claus Gernhardt Patentanwälte
Clemensstrasse 30 Postfach 40 14 68
W-8000 München 40 (DE)

(54) Apparatus for extracting facial image characteristic points.

(57) A facial image picture data is processed by an edge extraction part 2, which performs edge-stressing on the inputted image picture. And then the processed data is converted by a binary level conversion part 3 for each preestimated point of facial elements. The binary converted data is stored in a shape data-base part 4. The stored shape data is updated by an updating part 6. An image picture arithmetic processing part 5 performs time-to-time arithmetic computation of correspondence factors between the data of the binary-leveled edged image picture and the data stored in the shape data-base part 4, and issues an output including extracted data of image characteristic points.

FIG. 1



EP 0 552 770 A2

FIELD OF THE INVENTION AND RELATED ART STATEMENT**1. FIELD OF THE INVENTION**

5 The present invention relates to a characteristic points extraction part of an individual person identification apparatus and an facial shape characteristics. It is exemplified by a facial expression recognition apparatus for the facial image picture communication.

2. DESCRIPTION OF THE RELATED ART

10

Heretofore, as the method through which characteristic points or characteristic quantities are extracted from a facial image picture, there were, for example, Japanese unexamined Patent Publication (Tokkai) Sho-61-208185 (208185/1986) and Japanese unexamined Patent Publication (Tokkai) Sho-63-223974 (223974/1988). The former is such that a facial image picture inputted from a image picture input part is
 15 once stored in an image picture memory part, then it is binary-leveled in a binary level conversion part with an appropriate threshold value θ . Then the characteristic regions of a human face are extracted using characteristic parameters (area, circumferential length, coordinate of the center of gravity, etc.) from this binary-leveled image picture. The latter is such that the tint conversion is applied to a facial image picture stored in a image picture memory part, and then the obtained skin-color region is taken as a mask pattern
 20 and the characteristic regions are extracted.

In the above-mentioned prior art, in case that a binary-leveled image picture is used, when the threshold value θ varies, areas of obtained characteristic regions vary also. Hence there was a drawback that the characteristic points exhibited deviations depending upon the threshold value θ . Moreover, even for the same face, when the position of lighting source differs at the time of inputting the image picture, the
 25 brightness distribution on the face becomes different. Hence when the binary level conversion is done with a certain fixed threshold value of θ , areas of characteristic regions change. This causes the consequence that different extraction results are issued from a single same face.

Furthermore, when the hue is used as has been described above, there has been a problem that, depending on the sort of lighting sources, such as sun light, fluorescent lamp and others, the hue of
 30 respective regions including the skin-color regions change. Because of these, in the prior art, it was necessary to fix the position of lighting, colors, and others.

Also, in case that the hue information is used, if a television camera used as for the image picture input part, at the time of A/D-converting and storing the input signal into an image picture memory part, the hue information becomes unstable in such regions that is including sharp edges, making an accurate extraction
 35 of the characteristic points impossible.

OBJECT AND SUMMARY OF THE INVENTION

According to the present invention, in order to solve the above-mentioned problem, the characteristic
 40 points are extracted using the edged image picture, which is relatively stable even when the state of the light source varies. In order to achieve this, an apparatus of the present invention has an edge extraction part, and a binary level conversion part is provided for removing the noise of the edged image picture. And further, in order to catch the characteristic regions of the face such as mouth, eyebrow and others as their shape, a shape data-base part, in which the shapes thereof are stored, is provided. Furthermore, the
 45 apparatus is provided with a shape data updating part for updating the shape data in order to absorb the difference of faces among individuals and to fit the shape data to the inputted facial image picture.

In the present invention, for the facial image picture inputted from a image picture input part, an edged image picture is produced by the edge extraction part. The edged image picture includes a large amount of minute noise due to such as moustache or wrinkles. Therefore, for a searching region, i.e., a region to be
 50 searched in this image picture, the above-mentioned edged image picture thus obtained by the edge extraction part is converted into a binary-leveled edged image picture by a binary level conversion part. From the searching regions of obtained binary-leveled edged image picture, such a region that is close to the shape data stored in a shape data-base part is selected based on the magnitude of their correspondence factor obtained by an image picture arithmetic processing part. The shape data are updated in a
 55 manner that the correspondence factor becomes large in the vicinity of selected region by a shape data updating part. Then, when the correspondence factor outputted from the image picture arithmetic processing part reaches a certain value or more based on those updated shape data, the characteristic points that is the object of the search are outputted from the output part.

The apparatus of the present invention, owing to the binary level conversion of the edged image picture, is robust against the conditions for taking pictures, such as position of lighting source, color, and others. And since the shape data are stored as the data-base, even for the facial image picture wearing the glasses, for example, erroneous action of the apparatus becomes seldom. Furthermore, owing to the inclusion of the shape data updating part in the apparatus, personal difference depending on individual persons can be absorbed, enabling us to raise the capability of the characteristic points extraction.

While the novel features of the invention are set forth particularly in the appended claims, the invention, both as to organization and content, will be better understood and appreciated, along with other objects and features thereof, from the following detailed description taken in conjunction with the drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG.1 is a drawing of constitution of a first embodiment of the present invention.

FIG.2 shows an example of the input image picture and its searching regions.

FIG.3 shows an example of the shape data in the shape data-base part.

FIG.4 shows an example of shapes of respective facial elements.

FIG.5 shows a constitutional drawing of one embodiment of the present invention stated in claim 2.

FIG.6 shows a constitutional drawing of one embodiment of the present invention stated in claim 4.

FIG.7 shows a constitutional drawing of one embodiment of the present invention stated in claim 6.

FIG.8 shows a conventional drawing representing an example of hardware configurations of the present invention.

FIG.9(a) and FIG.9(b) in combination show a flow chart of an example of the procedure of extraction of the facial image characteristic points in the present invention.

It will be recognized that some or all of the Figures are schematic representations for purposes of illustration and do not necessarily depict the actual relative sizes or locations of the elements shown.

DESCRIPTION OF THE PREFERRED EMBODIMENT

In FIG.1, a constitutional drawing of a first embodiment of the present invention is shown. The output of the image picture input part 1 to which the facial images from a television camera or the likes are inputted is given to an edge extraction part 2, wherein the edge processing is applied to the inputted image picture. The output whereon the edge processing has been applied in the edge extraction part 2 is given to a binary level conversion part 3. In this part, for image picture data, binary level conversion process is performed for each preestimated region of respective facial elements. In a shape data-base part 4, shape data of facial elements such as iris, nose, mouth, and eyebrow are stored. Binary-leveled edged image picture, which is processed by the two-level conversion part 3, and the shape data from the shape data-base part 4 are inputted into the image picture arithmetic processing means 5, wherein the correspondence factor therebetween is computed. Starting from the correspondence factor thus obtained, contents of the shape data-base part 4 is updated in a manner that the corresponding factor increases by the shape data updating part 6. From the shape data of the shape data-base part 4 and the correspondence factor obtained by the image picture arithmetic processing part 5, the characteristic points of the image picture are extracted and issued from the output part 7.

In the following, explanation is given more to the details on the procedure of extracting the characteristic points.

First, the facial image picture is taken from the image picture input part 1, and an edged image picture is produced by the edge extraction part 2 from the inputted image picture. In this part, computing is made by using an operator such as, for example, the Sobel operator (see, for example, p.98 of D. H. Ballard and C. M. Brown, translated by Akio Soemura "Computer Vision", Japan Computer Association, 1987), wherefrom gradient vectors at respective pixels can be obtained. The gradient vectors thus obtained have their respective magnitudes as well as their directions. The direction of the gradient vector means a direction in which the gradient of brightness of the image picture takes a largest value, and the magnitude thereof means this largest value. Hereinafter, they are also called as the edge vectors, since those regions along such the pixels having large gradient vector magnitudes may form edges in the image picture. FIG.2 shows an example of the inputted image picture and the searching regions of respective facial elements.

In the following, taking the iris as an example, the procedure of extracting the characteristic points is described. First, the magnitudes m of edge vectors in a searching region of the iris shown in FIG.2 are converted into binary-leveled values of 0 or 1 using a certain threshold value θ . That is, the raw gradient vectors obtained by applying a gradient operation, such as Sobel operation described above, on the

brightness data of respective pixels (or positions), are normalized and converted into either of unit vectors or zero vectors. Hereinafter, for the convenience of explanation, sign of these unit vectors obtained as has been described above is reversed (multiplied by -1), and they are called again as the edge vectors (More accurately, they should be called as the normalized edge vector). This process is carried out in the binary level conversion part 3. Since the magnitudes m of edge vectors varies depending on the state of the lighting source used at the time of taking the image picture, the above-mentioned threshold value θ must be determined from the frequency distribution of the magnitude m . For example, the threshold value θ is determined in a manner that the binary level conversion is made by setting those data, which fall within 20 % probability of distribution from the largest magnitude in a relevant searching region, to be 1; whereas the rest those data falling within 80 % probability thereof to be 0.

In FIG.3 and FIG.4, an example of the contents stored in the shape data-base part 4 is shown. FIG.3 shows a shape data of an iris, as an example. In this case, number n of data-elements of the shape data is 12, that is, $n=12$. The shape data comprise 12 coordinate data and 12 gradient vectors at those respective coordinates. Coordinate data, l_k and m_k are coordinates at which gradient vectors, v_x and v_y are given. The gradient vector, v_y , are unit vectors giving the direction in which largest gradient value is present. Since the iris has a circular shape inside of which is much darker than the outside thereof, coordinate data form a circle and all the gradient vectors given at these coordinates direct to the center of the circle.

FIG.4 shows examples of facial elements and their shapes. Facial elements are iris, mouth, nose, eyebrow, and cheek, in this embodiment.

Next, the searching region for the edged image picture of iris is scanned and the correspondence factor ϕ between the inputted observed edge vectors and the gradient vectors of the shape data stored in the data-base is calculated in the arithmetic processing part 5. Since both of the inputted observed data and the shape data stored in the data-base part 4 are given in the form of vector, the correspondence factor ϕ to be required to calculate can be expressed by the average of inner products between those corresponding two vectors in a manner shown below.

Letting the inputted observed edge vectors in the edged image picture be

$$u_{ij} = (u_x, u_y)$$

where i, j are x, y coordinates of positions on the image picture and

$$u_x^2 + u_y^2 = 1,$$

and the shape data be

coordinate data: $p_k = (l_k, m_k)$

where l_k, m_k are x, y coordinates of positions whereat the shape data is given, respectively, and

gradient vectors: $v_k = (v_x, v_y)$

where $u_x^2 + u_y^2 = 1$ and

$1 \leq k \leq n$ (n is number of the positions),

then the correspondence factor ϕ of the shape data at coordinates (i,j) in the image picture can be expressed as

$$\phi = (\sum u_{i+l_k, j+m_k} \cdot v_k) / n,$$

where

$1 \leq k \leq n$ and

50

$$u_{i+l_k, j+m_k} \cdot v_k = u_x \cdot v_x + u_y \cdot v_y.$$

In a manner described as above and scanning the coordinate (i,j) , the correspondence factor ϕ for respective coordinates (i,j) in the searching region is calculated. Among them, a plural number of coordinates at which values of the correspondence factor ϕ are large are assigned to be preestimated regions of the relevant facial element.

Next, in respective preestimated regions, the shape data are updated by the shape data updating part 6, and then the correspondence factor ϕ is again searched. The scheme of updating is, for example, to move the coordinate of one position of the present data by +1 and then -1 in the direction of the gradient vector and to take either one direction in which the correspondence factor increases. After this movement, the direction of the gradient vector is also updated in a manner that it coincides with the shape data. In such the manner, all of the elements of shape data are successively updated in a manner that the correspondence factor ϕ is further improved. When the improvement of the correspondence factor ϕ stops, or when ϕ exceeds a certain specified value s , any further updating of the shape data is stopped.

Thereafter, in accordance with the updated shape data, the characteristic points are issued from the output part 7. The scheme of this outputting is as follows: For example, when a final values of the correspondence factor ϕ is less than a certain value t ($s > t$), regarding the region in which the correspondence factor ϕ is maximum to be a preestimated region and taking the shape data there to be a facial element to seek, only the necessary characteristic points thereof are issued. And in case that there are a plural number of preestimated regions in which ϕ is larger than a certain value t , the preestimated region is determined by, for example, a statistical procedure. That is, those regions which are disposed mutually close are all regarded to be genuine shape data. And then, by calculating the average with regard to corresponding positions of all of these shape data, new shape data are obtained.

Then, taking the obtained shape data to be the shape data of a facial element, namely the object to search for, only the necessary characteristic points are outputted. For example, in the case of iris, the average of coordinates of all the positions of the shape data are calculated to be a center point, and the maximum point and minimum point in the y-coordinate are taken to be the top and the bottom points of the iris, and then resultant data are issued from an output part.

Hereupon, since two irises are present in the x-coordinate direction on a face, it is necessary to issue two positions separated by a distance d or more at which the correspondence factor ϕ is large to be the two irises.

In a similar manner, the respective characteristic points of mouth, nose, eyebrow, and cheek can be extracted. For example, for the mouth, five points of top, bottom, left, right, and center are extracted; and for eyebrow, four points of top, bottom, left, and right are extracted.

And, in FIG.5, the constitution of a second embodiment is shown. In FIG.2, first, the searching region is selected only to the iris region, for example. And then, the searching regions for remaining facial elements are determined by the region determination part 8 based on the extracted characteristic points of the irises. In accordance with such process, it becomes possible to extract the characteristic points for remaining facial elements with a lesser amount of calculation. For example, the determination of the searching regions for the remaining facial elements can be processed by utilizing simple common knowledge such that the nose is present between mouth and eyes, and eyebrows are present immediately above eyes.

Furthermore, once the coordinates of two irises are determined, any possible tilt angle of the inputted facial image picture can be obtained. Hence, based on this tilt angle, by reversely rotating the shape data stored in the shape data-base part 4 by an amount of this obtained tilt angle by the shape data updating part 6, even from a tilted facial image picture, the extraction of the characteristic points becomes possible.

In FIG.6, a constitutional drawing of a third embodiment of the present invention is shown. The eyebrow has an edge which is not sharp but gradual. This is because the borders of hair of the eyebrow are gradual. Therefore, differing from other facial elements, for the eyebrow, it is difficult to obtain strong edge components. Consequently, for the extraction of the characteristic points of the eyebrow, by applying a preprocessing of binary level conversion on the searching regions of eyebrows by a binary level conversion part 11, it becomes possible to obtain strong edge components. This preprocessing is selected by the process selection part 10.

The application of the above capability of the present invention is not limited to the eyebrow, but also valid, for example, to such one as moustache wherein its edge component is also gradual. And, in particular, in case of extracting the characteristic points of the eyebrow, since the eyebrow is oblong horizontally, its brightness distribution differs largely between both ends. Consequently, if the searching region is binary-leveled at only one time, it can happen that an accurate shape of eyebrow does not appear. Then, (as in the aspect described in claim 5,) the searching region of the eyebrow is divided into small sub-regions in the vertical direction. In respective small sub-regions, respective threshold values for binary level conversion are respectively determined in a manner that j % probabilities of brightness distribution is set to 0. Hereupon, j is determined in accordance with, for example, the area of respective searching regions. By obtaining the average and variance of the threshold values for respective small sub-regions, respective regions can be binary-leveled individually. At this process, in case that the threshold value deviates largely from the average value, it is regarded to be either one of

(i) There is no eyebrow in that small sub-region,
or

(ii) There exist a lot of hair. Thus the process is carried out on respective small sub-regions by which the whole parts of those small sub-regions are binary-leveled totally to 0 or 1 in accordance with the above classifications.

The present embodiment is valid not only for such the case that the lighting source deviates from the center of the face to the right or left direction, but also it is effective in that the influence of hair existing in the eyebrow region can be reduced.

In FIG.7, a constitutional drawing of a fourth embodiment of the present invention is shown. In the case of the present embodiment, the preestimated regions obtained for respective facial elements are recorded in the adoption/rejection part 9; and from the combination thereof, one giving an adequate facial shape is selected. As the condition to obtain an adequate facial shape, the following facts, for example, can be used:

(iii) Nose and mouth are present on an equidivision-perpendicular line between two irises,

(iv) Distance between two eyebrows and that between two irises are almost the same, and

(v) The right cheek and the left cheek are present at almost equidistant places to the right hand and the left hand from the above equidivision-perpendicular line.

In such the manner, the best fit preestimated regions are searched for respective facial elements. And then for respective preestimated regions, the characteristic points which are the object of search can be obtained based on the shape data.

In the above-mentioned example, explanation of the present invention has been given on four different embodiments. Such embodiments may be constituted by hard ware as shown in the Figures, but may be constituted also by data processor either. In order to elucidate the present invention embodied by using the data processor or a computer specifically and concretely, an example of hardware configuration of computer for use in the present invention is shown in FIG.8, and an example of the procedure of extraction of the facial image characteristic points, which has been already described in the above embodiments, is now explained using a flow chart shown in FIG.9(a) and FIG.9(b).

FIG.8, shows a circuit block diagram giving a fundamental hardware configuration of the apparatus of the present invention. The facial image picture is taken into the apparatus through a television camera 101. The facial image picture signal issued from the television camera 101 is inputted into an A/D converter 102. A central processing unit, CPU 103 executes all the required functions, such as data access, transfer, store, arithmetic processing, and other functions for data under instructions of program installed in the apparatus. Functions or parts represented by boxes in FIG.5 through FIG.7 are preferably executed by such the installed program. Numeral 104 designates an image picture memory. The output of the A/D converter is memorized through a CPU 103 in an input image picture memory 104A as the input image picture data for all of each pixel. The input image picture data are converted into edged image picture data and further converted binary-leveled edged image picture data through the CPU 103. They are stored in an edged image picture memory 104B and a binary-leveled edged image picture memory 104C, respectively. Numeral 105 is a memory for storing the shape data-base of facial elements such as eyes, mouth, eyebrow, or cheek. data-base of each facial element includes three different sizes of small, medium and large. The correspondence factor between the binary-leveled edged image picture data and the shape data stored in the shape data-base memory 105 is computed; and the shape data are updated in a manner that the correspondence factor increases by the CPU 103. Numeral 106 is a working area of memory used for temporary purpose of the processing. Numeral 107 is an output area of memory for storing the extracted facial image characteristic points of necessary facial elements.

In FIG.9(a) and FIG.9(b) in combination, a flow chart of an example of the procedure of extraction of the facial image characteristic points is shown. In FIG.9(a), a flow starting at a start 201 through a step 216 corresponds to the process for the extraction of facial image characteristic points of eyes, whereas in FIG.9-(b), flow of a step 302 through a step 316 corresponds to the process for the extraction of facial image characteristic points of mouth. For remaining other facial elements, almost the same flow chart as for the above two facial elements can be applied.

Hereupon, in the present invention it is unnecessary to use any color picture image, and thereby the extraction of the characteristic points is possible even from a monochromatic photograph. For the shape data, by preparing a plural number of data for one facial element, the obtainable accuracy of extraction of the characteristic points can be improved.

Although the present invention has been described in terms of the presently preferred embodiments, it is to be understood that such disclosure is not to be interpreted as limiting. Various alterations and modifications will no doubt become apparent to those skilled in the art after having read the above

disclosure. Accordingly, it is intended that the appended claims be interpreted as covering all alterations and modifications as fall within the true spirit and scope of the invention.

Claims

5

1. An apparatus for extracting facial image characteristic points comprising:
 - an image picture input part (1, 101, 102, 104A) where through the facial image data are inputted,
 - an edge extraction part (2, 103, 104B) for performing edge-stressing process on edge parts included in said facial image data to make an edged-image data,
 - 10 a binary level conversion part (3, 103, 104C) for performing conversion of the edged-image data into binary-leveled edged-image data on each preestimated region of facial elements,
 - a shape data-base part (4, 105) which stores a plurality of shape data as reference data-base of the facial elements,
 - a shape data updating part (6, 103) for updating the data in the shape data-base part,
 - 15 an image picture arithmetic processing part (5, 103, 106) for performing the arithmetic computation of the magnitudes of correspondence factors between the binary-leveled edged-image data and said shape data stored in the shape data-base part, and
 - an output part (7, 107) for issuing output data including an information of image characteristic points extracted based on said correspondence factors and said shape data.

20

2. An apparatus for extracting facial image characteristic points in accordance with claim 1 which further comprises:
 - a region determination part (8, 103) for fixing a searching region of a facial element wherefrom the first extraction process is performed, the region determination for rest facial elements being carried out based on the position information of the facial element which has been extracted previously.

25

3. An apparatus for extracting facial image characteristic points in accordance with claim 2 wherein in the characteristic points extraction process, positions of irises are taken to be an item wherefrom the first characteristic extraction process is carried out.

30

4. An apparatus for extracting facial image characteristic points in accordance with claim 2 which further comprises:
 - a process selection part (10, 103) and
 - a binary level conversion part (11, 103) for performing said binary level conversion process in advance to said edge extraction for particular facial elements.

35

5. An apparatus for extracting facial image characteristic points in accordance with claim 4 which further comprises:
 - a binary level conversion part (11, 103) in which a preestimated region, wherein the binary level conversion is to be processed, is divided into a plural number of smaller sub-regions, and binary level conversion threshold value for each smaller sub-region is determined from the brightness distribution of each smaller sub-region.

40

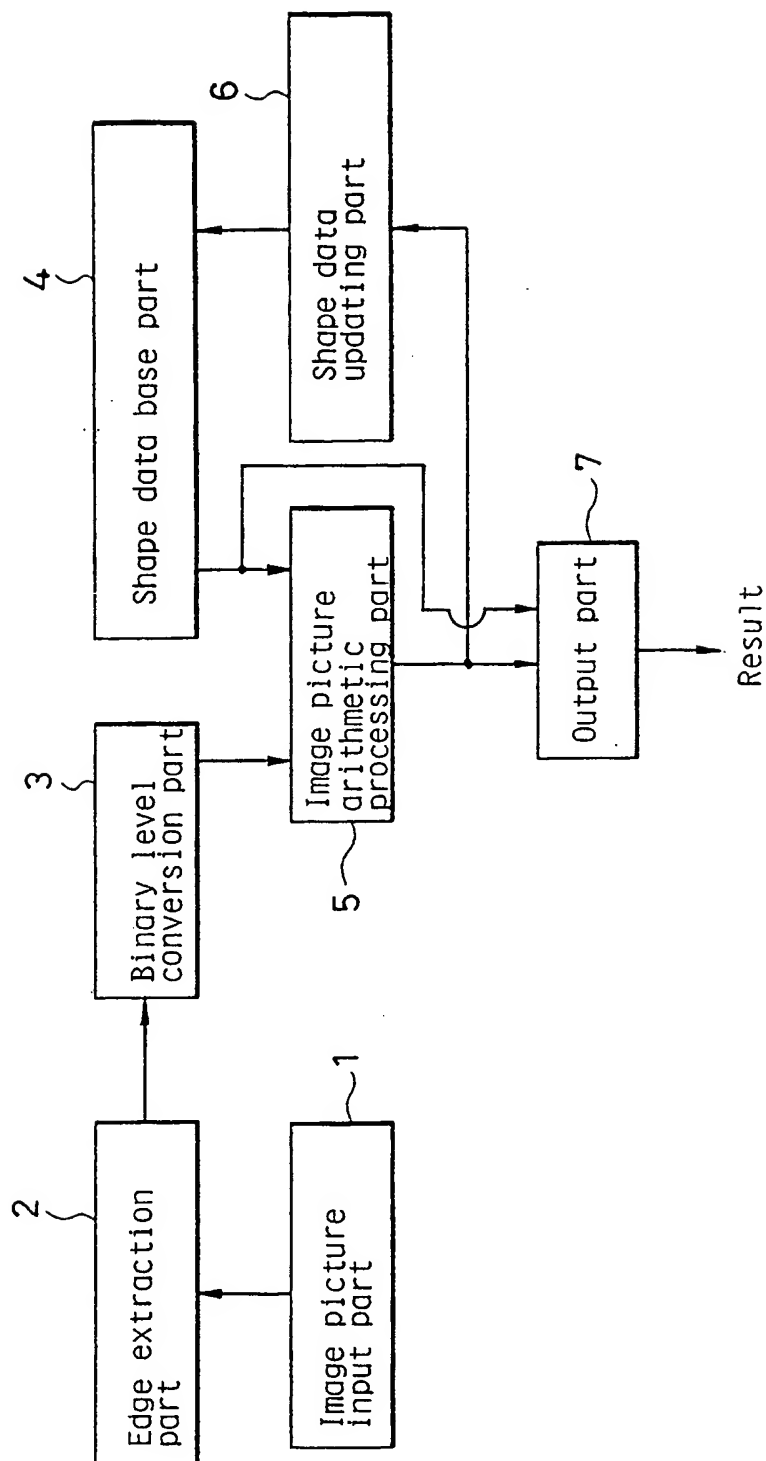
6. An apparatus for extracting facial image characteristic points in accordance with claim 1 which further comprises:
 - an adoption/rejection part (9, 103) whereby, after obtaining preexpected positions of all the facial elements, said image characteristic points of image picture are extracted by discarding those erroneously extracted facial elements using said adoption/rejection part.

45

50

55

FIG. 1



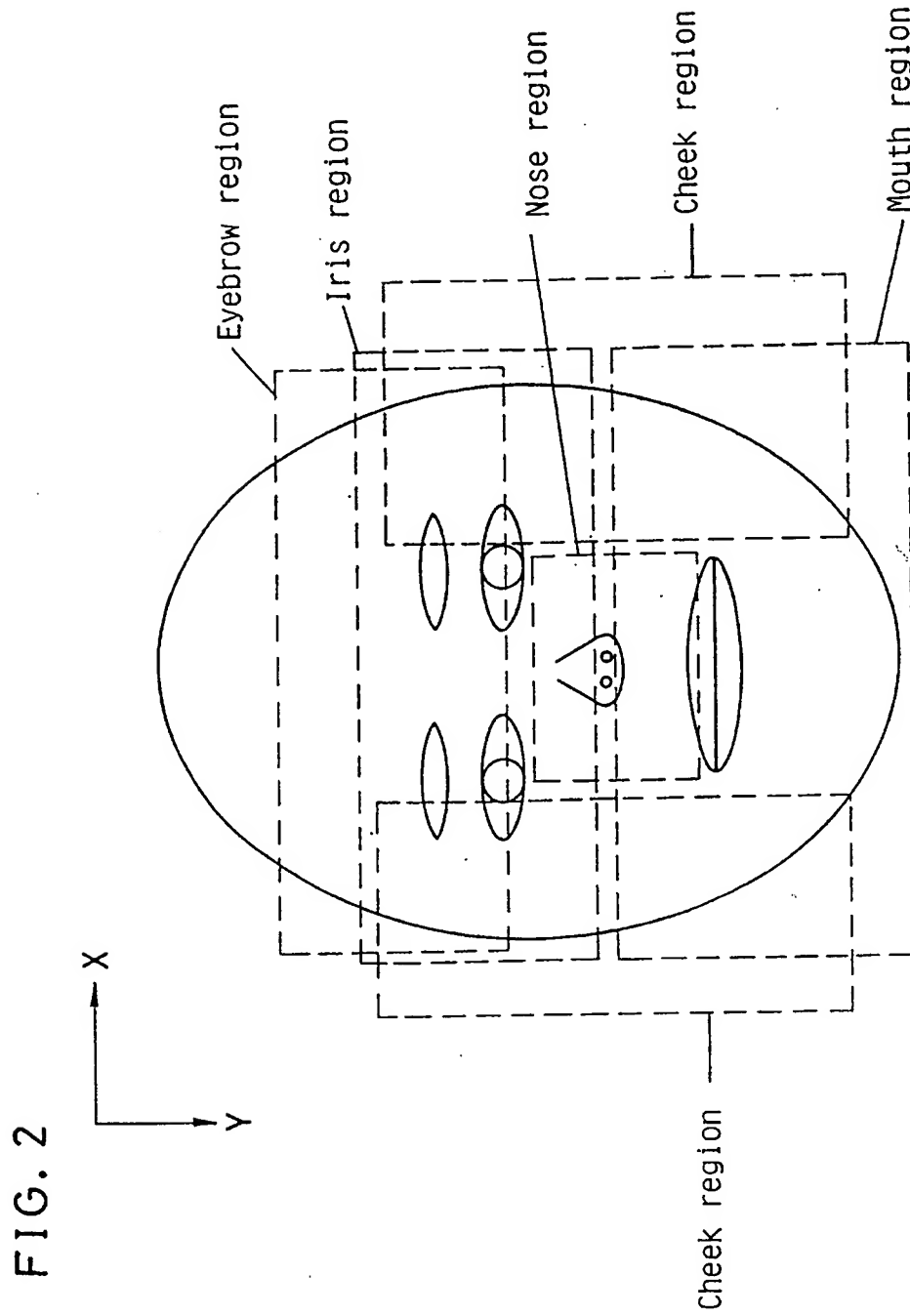


FIG.3

Position number k	l_k	m_k	V_x	V_y
1	6	1 2	0	- 1
2	9	1 1	- 0. 5	- 0. 8 6 6 1
3	1 1	9	- 0. 8 6 6 1	- 0. 5
4	1 2	6	- 1	0
5	1 1	3	- 0. 8 6 6 1	0. 5
6	9	1	- 0. 5	0. 8 6 6 1
7	6	0	0	1
8	3	1	0. 5	0. 8 6 6 1
9	1	3	0. 8 6 6 1	0. 5
1 0	0	6	1	0
1 1	1	9	0. 8 6 6 1	0. 5
1 2	3	1 1	0. 5	0. 8 6 6 1

FIG. 4

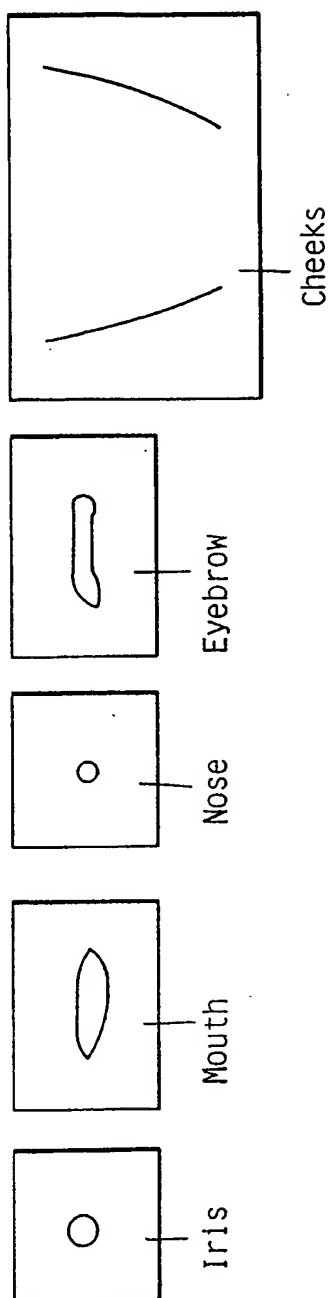


FIG. 5

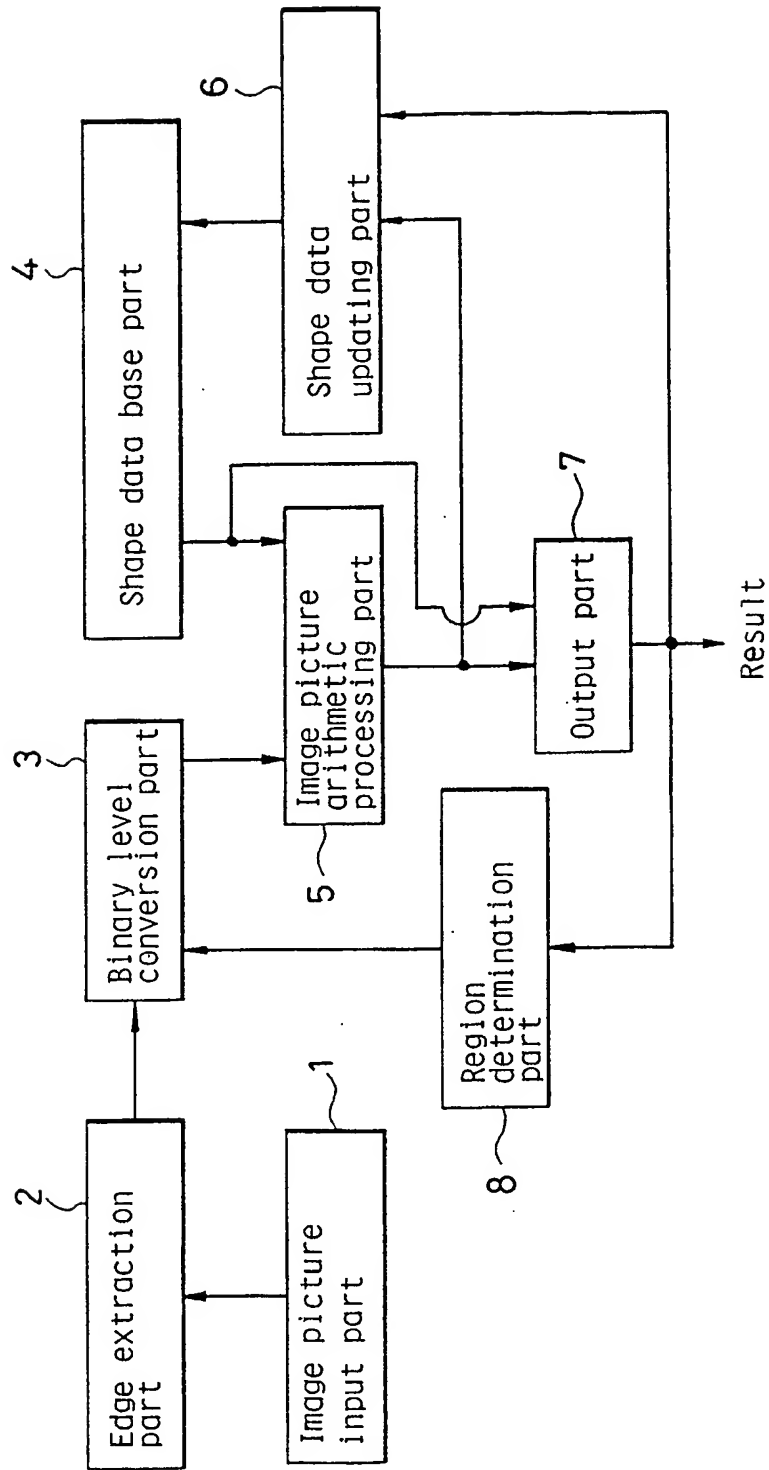
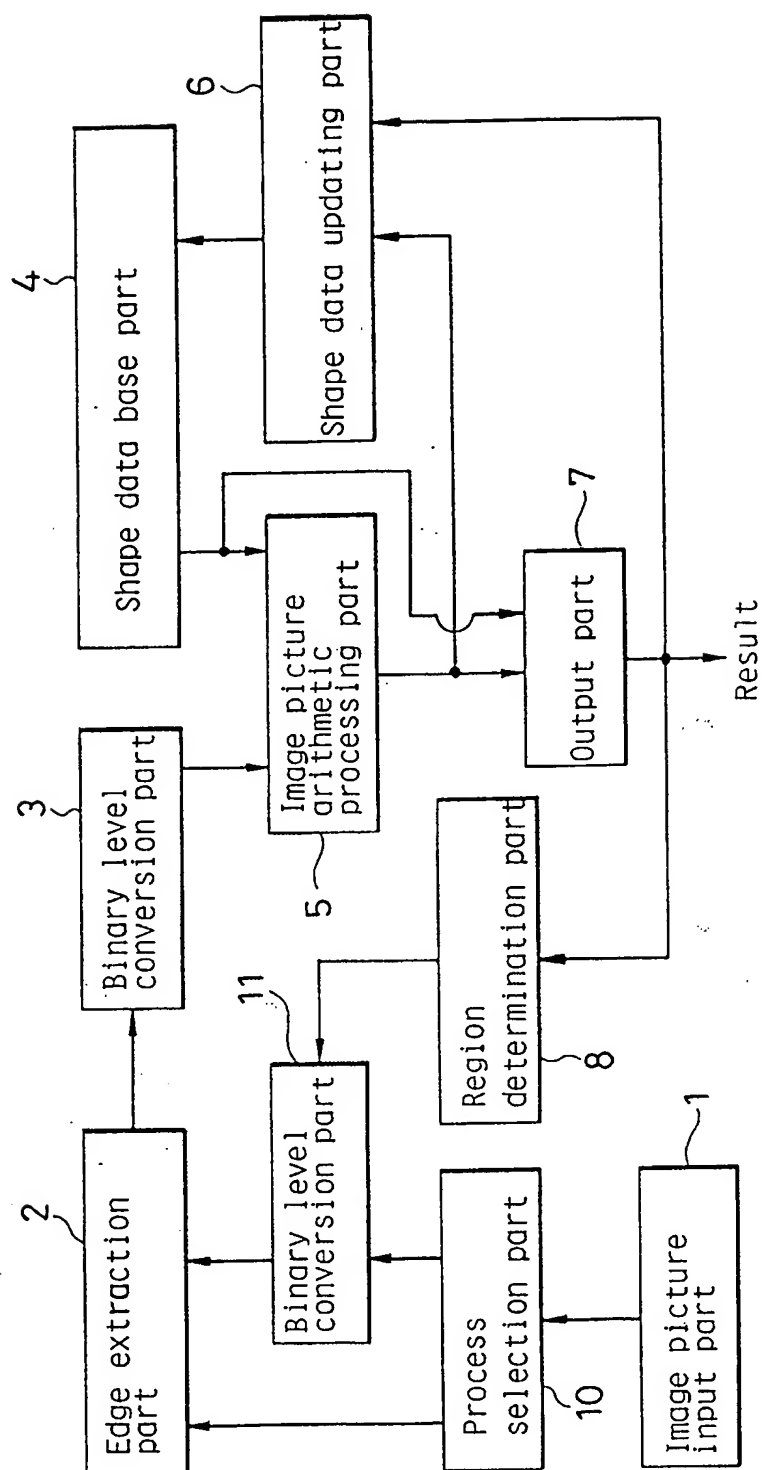


FIG. 6



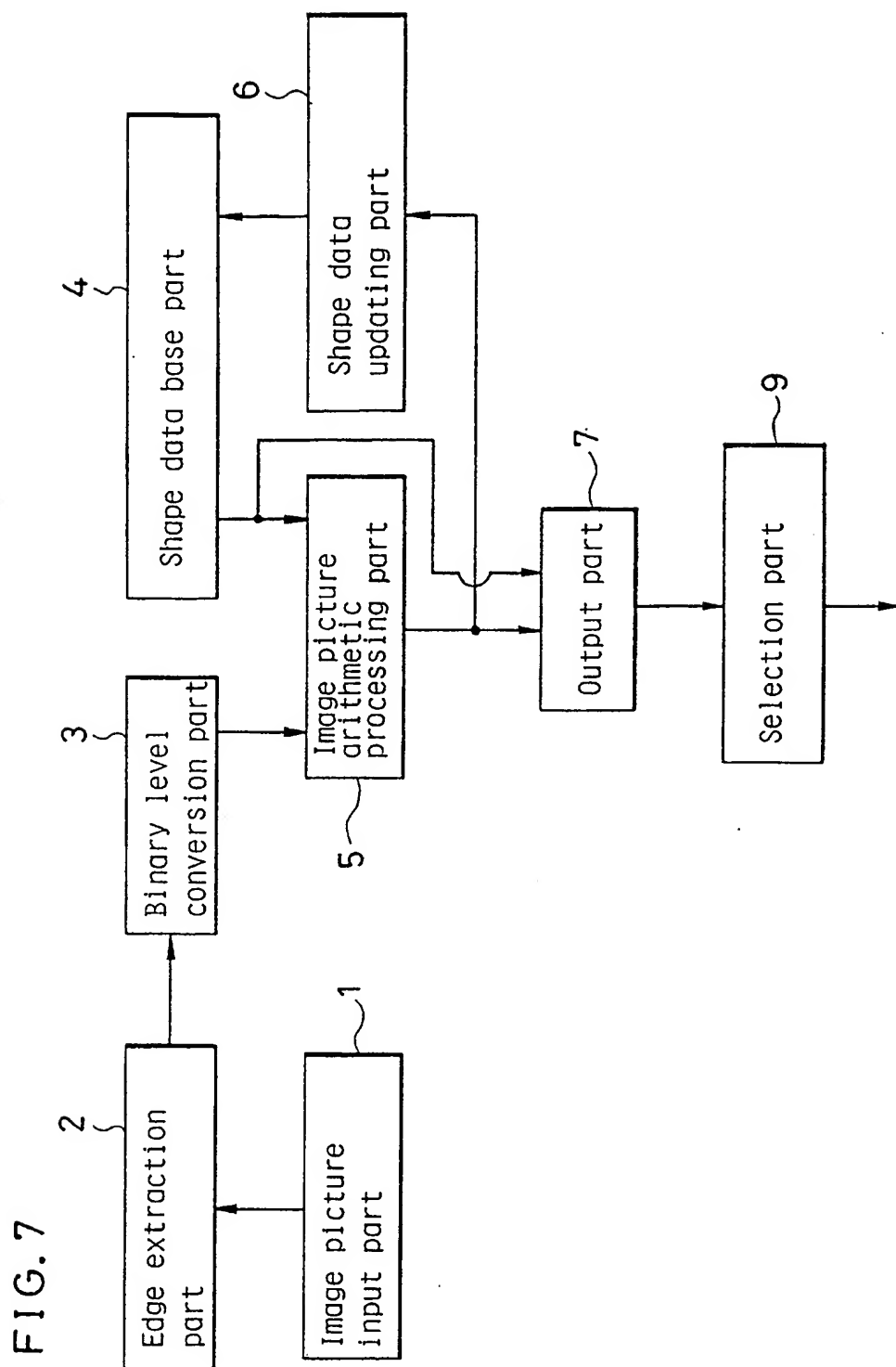


FIG. 8

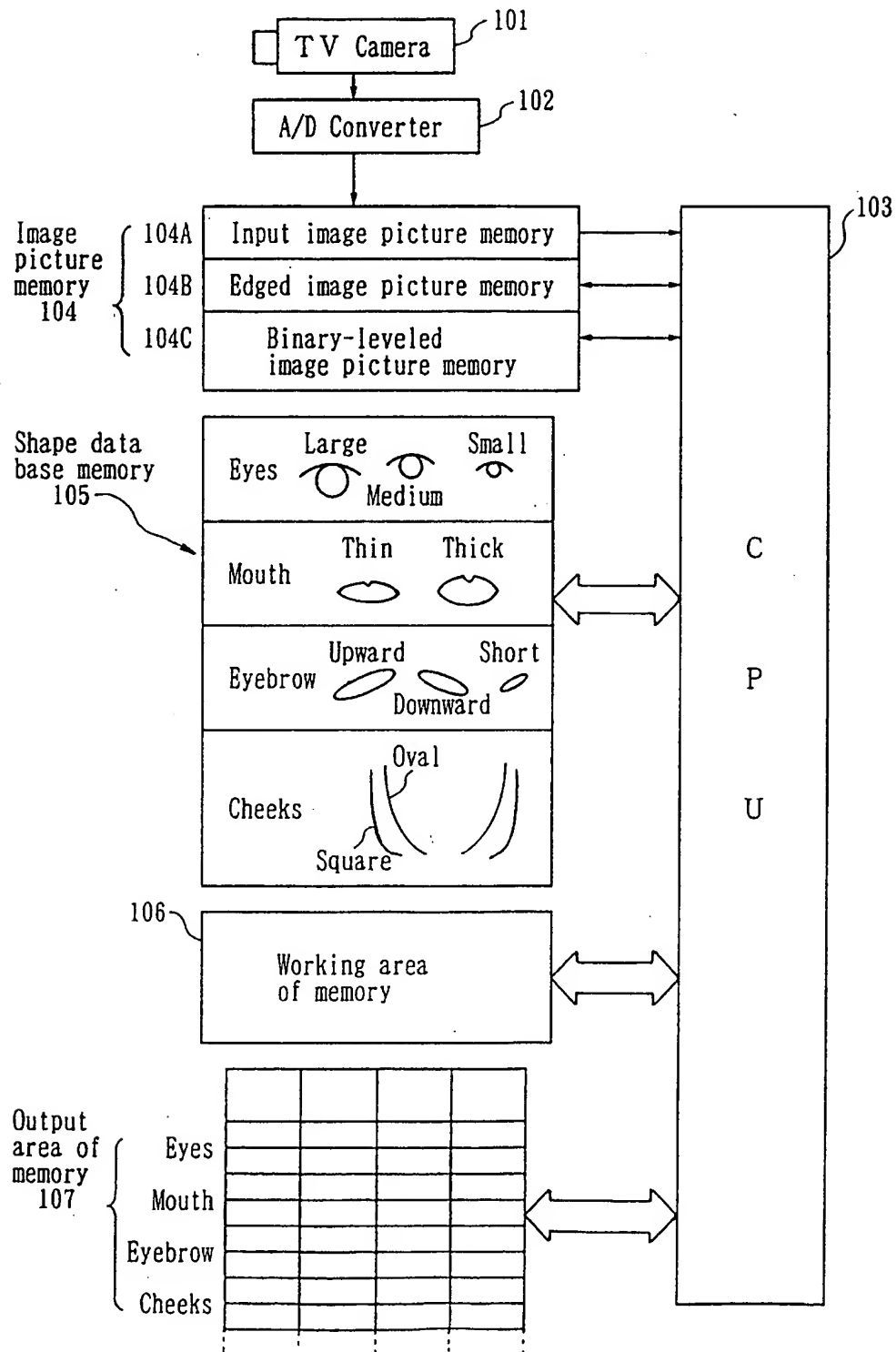


FIG. 9(a)

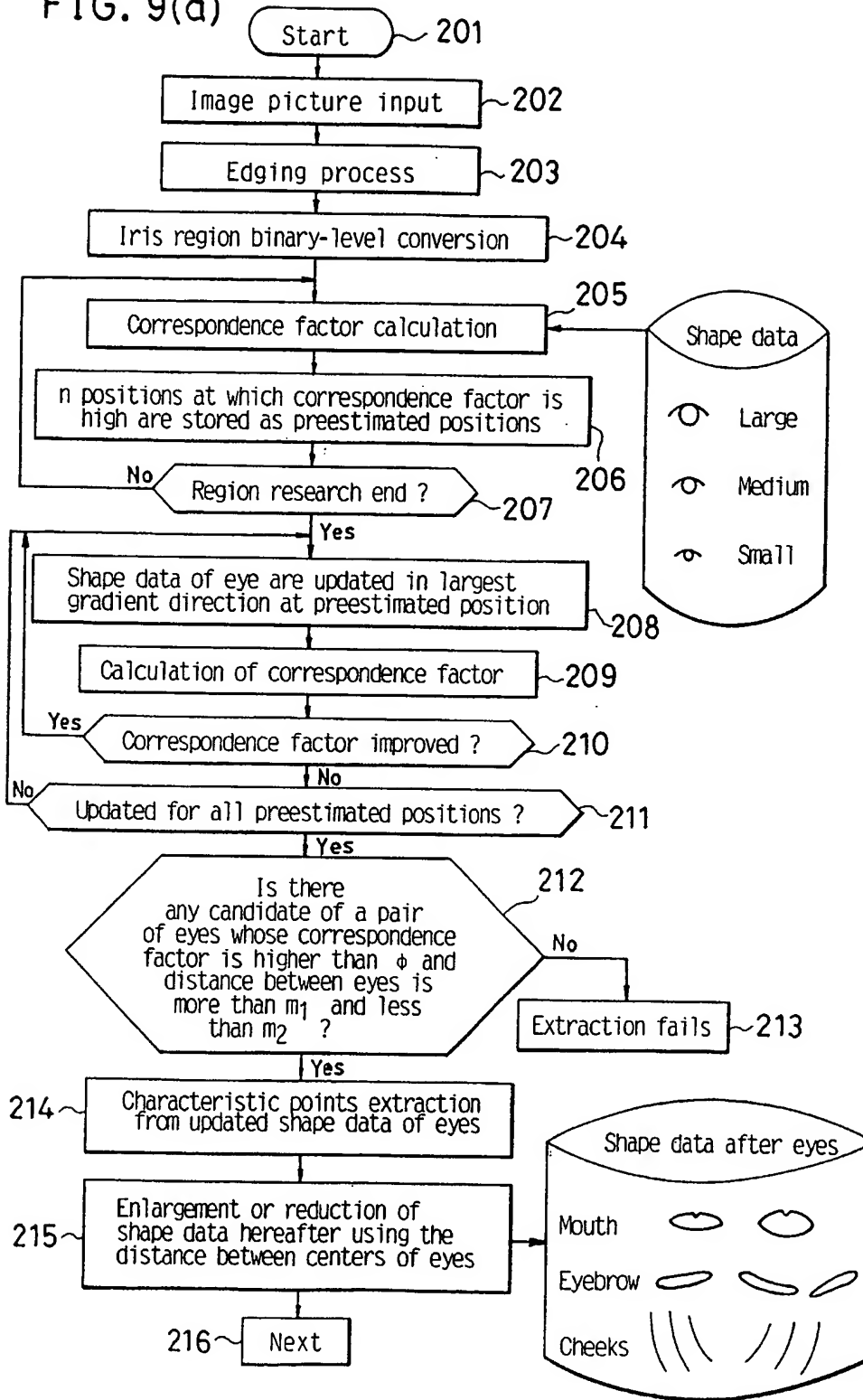
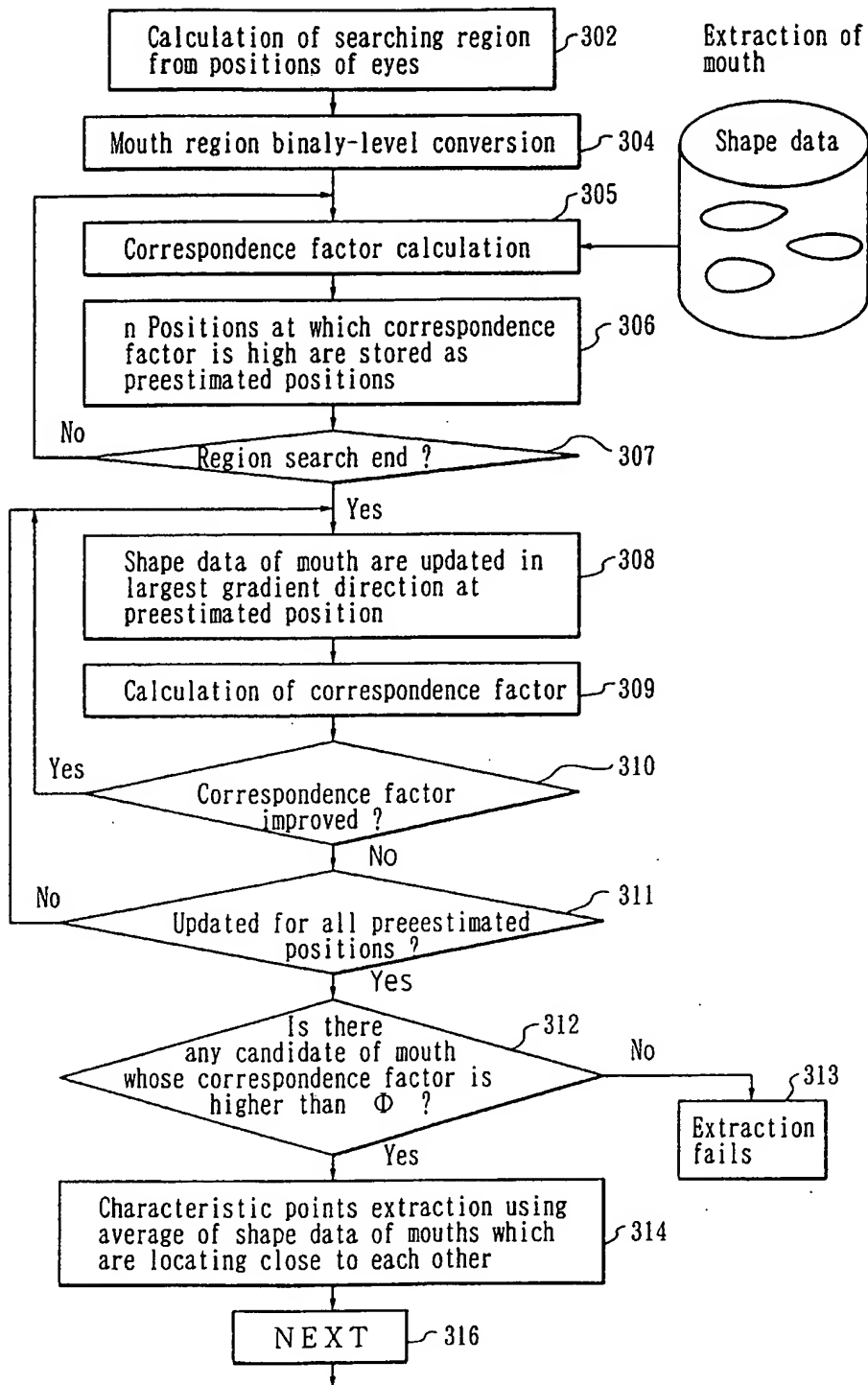


FIG. 9(b)



(Hereafter, Eyebrow and Cheeks are examined, and candidate close to the stored reference data are sought for.)

THIS PAGE BLANK (USPTO)